Sequential mapping – a novel approach to map from ICD-10-CM to ICD-11

Julia Xu, MD, PhD; Kin Wah Fung, MD, MS, MA; Olivier Bodenreider, MD, PhD

National Library of Medicine, National Institutes of Health, USA

Abstract

Background: ICD-11 will be used to report mortality statistics by WHO member countries starting in 2022. In the US, ICD-10-CM will likely continue to be used for morbidity coding for a long period of time. A map between ICD-10-CM and ICD-11 will therefore be useful for interoperability purpose between datasets coded with ICD-10-CM and ICD-11.

Objectives: The objective of this study is to explore novel approaches to automatically derive a map between ICD-10-CM and ICD-11 through the sequential use of existing maps.

Methods and results: Sequential mapping through ICD-10 yielded better coverage and accuracy compared to mapping through SNOMED CT.

Conclusions: Sequential mapping is useful in automatically creating a draft map from ICD-10-CM to ICD-11 and would reduce manual curation efforts in creating the final map. The various approaches offer different trade-offs among coverage, recall and precision.

Keywords:

ICD-10-CM; SNOMED CT; ICD-11

Introduction

The World Health Organization (WHO) adopted the 11th edition of the ICD (ICD-11) in May 2019. It is expected to be implemented by member countries starting in January 2022. [1-3] A significant upgrade from earlier revisions, ICD-11 will serve as the global standard for health data, clinical documentation, and statistical aggregation. Expected to be fully operating in an electronic environment, [4] ICD-11 was developed to accurately reflect contemporary medical practice and capture more information for morbidity use cases to improve quality of primary care and patient safety.

As a WHO member country, the US is required to report mortality statistics using the latest version of ICD. [5] For morbidity, the US has created a separate extension called Clinical Modification (ICD-9-CM and ICD-10-CM) to provide additional clinical details. The adoption of ICD-10-CM for morbidity coding in the US occurred in 2015, [6] 16 years after ICD-10 was released for mortality reporting. Therefore, there will likely be a considerable delay before ICD-11 is used for morbidity coding in the US. In other words, there will be a significant period of time (likely 10 years or more) in which ICD-10-CM and ICD-11 will coexist. (This is similar to the situation in which data coded with ICD-9-CM and ICD-10 coexisted.)

A map between ICD-10-CM and ICD-11 [7] will support interoperability among datasets coded with these two standards. More specifically, in countries such as the US, with claims data coded with ICD-10-CM and mortality data coded with ICD-11, the map will support the integration of mortality and morbidity data. Internationally, the fact that different countries may adopt ICD-11 for morbidity coding at different times will also create interoperability issues when aggregating morbidity datasets internationally. Here again, a map will support the integration of morbidity data coded with the two standards.

Creating a map between two coding standards is a challenging and labor-intensive process, because a high-quality map usually requires substantial manual curation. However, algorithmic mapping approaches can help to reduce the manual curation effort and even improve mapping accuracy. [8] The objective of this study is to explore novel approaches to automatically derive a map between ICD-10-CM and ICD-11 through the sequential use of existing maps. [9] More specifically, we investigate the following two sequences: ICD-10-CM → SNOMED CT → ICD-11 and ICD-10-CM → ICD-10 → ICD-11.

While combining existing maps to derive new maps is one of the techniques used for ontology alignment [10], for example, aligning anatomical ontologies, [11] we believe our study is the first report of using sequential mapping to map from ICD-10-CM to ICD-11. Additional contributions of this work include a comparison between two sequential mapping approaches and a careful failure analysis that provides insights about the benefits of this technique.

Materials

As the substrate for our sequential mapping approach, we used three existing, publicly available maps:

1) SNOMED CT to ICD-10-CM map (**NLM map**), a rule-based map published by the US National Library of Medicine (NLM) to support semi-automated generation of ICD-10-CM codes from clinical data encoded in SNOMED CT; 2) ICD-10 to ICD-11 map (**WHO map**), published by WHO and included in the ICD-11 implementation package; and 3) SNOMED CT to ICD-11 map (**SI map**), a prototype map published by SNOMED International using automatic mapping algorithms. Of note, this map has not been quality assured and is not intended to be used in production systems.
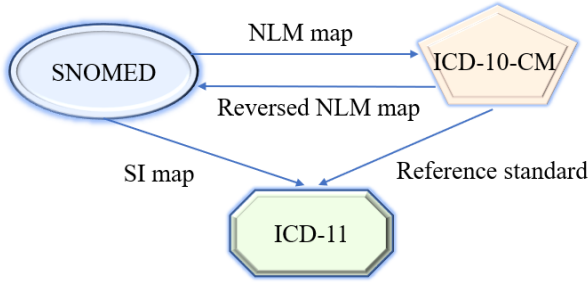
Methods

We explore two sequential mapping methods and their combinations, and we evaluate these methods against a reference mapping.

A. Mapping methods

1. ICD-10-CM → SNOMED CT → ICD-11 (**method 1**)

Figure 1. Method 1: Sequential mapping using NLM and SI maps

As shown in Figure 1, we first mapped ICD-10-CM codes to SNOMED CT concepts by using the NLM map. Note that the original direction of this map was from SNOMED CT to ICD-10-CM, but we used it in the reverse direction, which may lead to potential issues discussed in the error analysis. We then used the SI map to map from SNOMED CT to ICD-11. Since we used the NLM map to go from ICD-10-CM to the finer-grained SNOMED CT, we did the following to minimize meaning drift.

- Only default mapping targets were used, excluding the rule-based maps (e.g., age rules with different codes for different age groups)

- Sometimes a SNOMED CT concept was mapped to a combination of ICD-10-CM codes (e.g., ICD-10-CM required separate codes for the etiology and manifestation of a condition). In the NLM map, each ICD-10-CM code was put into a Mapgroup (e.g., a SNOMED CT concept with 3 ICD-10-CM target codes had 3 Mapgroups). Usually, Mapgroup 1 had the most relevant code.

We compared the results of using all Mapgroups and only Mapgroup 1.

- In the output of this step, if multiple SNOMED CT targets were found for the same ICD-10-CM code, we only kept the highest level SNOMED CT concept. For example, the ICD-10-CM code K74.60 *Unspecified cirrhosis of liver* was mapped to the SNOMED CT concept 103611000119102 *Cirrhosis of liver due to hepatitis B* and its ancestor 19943007 *Cirrhosis of liver*. We only kept the coarser 19943007 *Cirrhosis of liver*.

2. ICD-10-CM → ICD-10 → ICD-11 (**method 2**)

Figure 2. Method 2: Sequential mapping using code conversion and WHO map
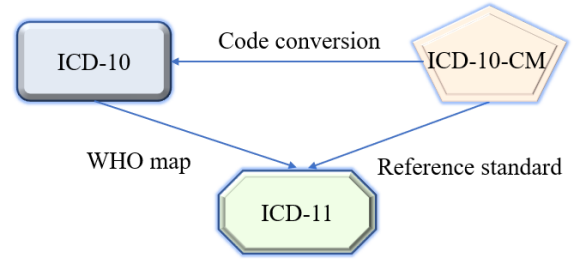
As shown in Figure 2, we first converted ICD-10-CM codes to ICD-10 codes, then used the WHO map to map from ICD-10 to ICD-11. Code conversion was done by truncating an ICD-10-CM code to arrive at an ICD-10 code that was included in the WHO map. The ICD-10 code could be a leaf code or higher level code. For example, the ICD-10-CM code S05.12XA was converted to the ICD-10 code S05.12 (leaf code) but S05.12 was not found in the WHO map. We further truncated the code to S05.1 which was covered by the WHO map. The WHO map came in two flavors: One-category map (one-to-one) and Multiple-category map (one-to-many). We used both and compared their performance.

Additionally, we considered the union and the intersection of the best-performing variant of methods 1 and 2.

B. Evaluation

To evaluate the various methods, we used a reference standard created in a previous study that involved 943 commonly used ICD-10-CM codes representing 60% of usage in large claims and hospital data sets (publication pending). The ICD-10-CM codes were independently mapped to ICD-11 by two terminologists and differences discussed until consensus was reached. We calculated the following statistics for each method: Coverage (% of ICD-10-CM codes for which any ICD-11 target could be found); Recall (% of correct ICD-11 targets in the reference standard that was identified); Precision (% of

correct ICD-11 targets among all ICD-11 targets found); and F-1 score (harmonic mean of recall and precision).

We also performed a failure analysis on methods 1 and 2.

Results

A. Overall performance

In the reference standard, 943 ICD-10-CM codes were each mapped to one ICD-11 code. The performance of the various methods is summarized in Table 1. For method 1, if we only used Mapgroup 1, the coverage dropped slightly, but the F-1 score was better. Similarly, for method 2, using the One-category map gave better F-1, with a slightly lower coverage. Among the combined methods, the union of method 1 (using Mapgroup 1 only) and method 2 (using One-category map) had better coverage and F-1 score compared to the intersection.

Table 1 – Sequential Mapping performance

	ICD-10-CM codes covered	Coverage	Recall	Precision	F-1 score
Method 1					
All map groups	681	72.2%	45.1%	25.4%	32.5%
Map group 1 only	671	71.2%	44.2%	28.2%	34.4%
Method 2					
Multiple Categories map	903	95.8%	55.8%	44.5%	49.5%
One Category map	894	94.8%	55.0%	58.1%	56.5%
Methods 1+2					
Union	929	98.5%	70.4%	32.7%	44.6%
Intersection	343	36.4%	29.2%	80.2%	42.8%

Table 2 – Failure analysis for method 1 (Using Mapgroup 1 only)

	Failure in step 1	Failure in step 2	Error in reference standard
# of unique ICD-10-CM codes	176 (61.1%)	108 (37.5%)	4 (1.4%)
# of map records	671 (79.4%)	170 (20.1%)	4 (0.5%)

B. Failure analysis

We reviewed all cases in which the ICD-11 target found was different from the reference standard.

1. Failure analysis for method 1 (Mapgroup 1 only)

i) Failure in step 1 (ICD-10-CM → SNOMED CT)

Of the 845 incorrect map records (pairs of ICD-10-CM and ICD-11 codes), 671 (79.4%) involving 176 unique ICD-10-CM codes mapped to the wrong ICD-11 target because of problems in the first step of the sequential map. There were three types of problems.

a. ICD-10-CM code mapped to an overly specific SNOMED CT concept

For example, ICD-10-CM code J44.9 *Chronic obstructive pulmonary disease, unspecified* was mapped to the more specific SNOMED CT concept 40100001 *Obliterative bronchiolitis* through the NLM map (which was used in the reverse direction). In the subsequent step, this SNOMED CT concept was mapped to the ICD-11 code CA26.0 *Chronic obliterative bronchiolitis*. The correct ICD-11 target for the ICD-10-CM code J44.9 should be CA22.Z *Chronic obstructive pulmonary disease, unspecified*.

b. ICD-10-CM code mapped to an overly broad SNOMED CT concept

For example, the ICD-10-CM code F32.9 *Major depressive disorder, single episode, unspecified* was mapped to the SNOMED CT concept 35489007 *Depressive disorder*. This SNOMED CT concept was subsequently mapped to the ICD-11 code 6A7Z *Depressive disorders, unspecified*. As a result, the more appropriate ICD-11 target 6A70.3 *Single episode depressive disorder, severe, without psychotic symptoms* was missed.

c. ICD-10-CM code mapped to a composite SNOMED CT concept

For example, the ICD-10-CM code D64.9 *Anemia, unspecified* was mapped to a composite SNOMED CT concept 43742007 *Pericarditis associated with severe chronic anemia*, which was subsequently mapped to the ICD-11 code BB22 *Constrictive pericarditi*s, an obviously wrong target.

ii) Failure in step 2 (SNOMED CT → ICD-11)

Of the 845 incorrect map records (pairs of ICD-10-CM and ICD-11 codes), 170 (20.1%), involving 108 unique ICD-10-CM codes, failed due to the errors in the SI map. For example, the SNOMED CT concept 122409911000119102 *Squamous cell carcinoma of right lung* was incorrectly mapped to the ICD-11 code 2B60.1 *Squamous cell carcinoma of lip* in the SI map.

iii) Error in reference standard

In four cases, the error was actually in the reference standard. For example, in the reference standard, the ICD-10-CM code *C18.7 Malignant neoplasm of sigmoid colon* was mapped to the ICD-11 code 2B90.3 *Malignant neoplasm of sigmoid colon* which was a higher level (non-leaf) code not valid for coding. Instead, the descendant 2B90.3Z *Malignant neoplasm of sigmoid colon, unspecified* should be used.

2. Failure analysis for method 2 (using One-category map)

Table 3 – Failure analysis for method 2 (using One-Category map)

	Failure in step 1	Failure in step 2	Error in reference standard
# of unique ICD-10-CM code	188 (50.1%)	184 (49.1%)	3 (0.8%)
# of map** records	188 (50.1%)	184 (49.1%)	3 (0.8%)

**since method 2 always resulted in one-to-one mappings between ICD-10-CM and ICD-11, the unique ICD-10-CM code and map record counts were the same

i) Failure in step 1 (ICD-10-CM → ICD-10)

Due to the incomplete coverage of the WHO map for ICD-10 codes, in order to reach a target ICD-11 code, sometimes we had to select an ICD-10 code which was the parent of a valid ICD-10 code. We called this "code roll-up". As shown in Table 3, this was the cause of failure in 188 ICD-10-CM codes. For example, the ICD-10-CM code R35.0 *Frequency of micturition* was rolled up to R35 *Polyuria*. In the WHO map, R35 was mapped to the ICD-11 code MF55 *Polyuria*. The correct ICD-11 target for the ICD-10-CM code R35.0 should be MF50.0 *Frequent micturition*.

Overall, as shown in Table 4, when there was no need for code roll-up, the final ICD-11 map targets were more likely to be correct than when there was code roll-up (71.6% vs. 53.3%).

Table 4 – Effect of ICD-10 code roll-up on accuracy of ICD-11 map target

	ICD-11 target correct	ICD-11 target incorrect	Total
No code roll-up	167 (71.6%)	66 (28.3%)	233
Code roll-up	352 (53.3%)	309 (46.7%)	661

ii) Failure in step 2 (ICD-10 → ICD-11)

Of the 375 incorrect map records (pairs of ICD-10-CM and ICD-11 codes), 184 (49.1%) maps failed due to errors in the WHO map. Here are some examples:

- The ICD-10 code F05 *Delirium due to known physiological condition* was mapped to the ICD-11 code 6D70.Z *Delirium, unspecified or unknown cause*. Instead, 6D70.Y *Delirium, other specified cause* should be used.

- The ICD-10 code L30.9 *Dermatitis, unspecified* was mapped to the ICD-11 code EA89 *Generalised eczematous dermatitis of unspecified type*. In the ICD-11 index, "dermatitis" pointed to another code EA8Z *Dermatitis or eczema, unspecified*.

- The ICD-10 code K59.0 *Constipation* was mapped to the ICD-11 code DE2Z *Diseases of the digestive system, unspecified,* which should be ME05.0 *Constipation*.

iii) Error in reference standard

Three of the four cases mentioned above in method 1 failure analysis occurred in method 2 as well.

Discussion

Findings

We have explored two sequential mapping approaches to automate mapping from ICD-10-CM to ICD-11 using existing, publicly available maps. Our results show that method 2 (going through ICD-10) is better in terms of coverage and F-1 score.

In the failure analysis for method 1 (going through SNOMED CT), most of the errors result from the first step of going from ICD-10-CM to SNOMED CT. This is not surprising given that we are using the NLM map in the reverse of its intended direction. Since SNOMED CT is finer-grained than ICD-10-CM, going from ICD-10-CM to SNOMED CT will give rise to a lot of one-to-many matches, causing meaning drift and loss of accuracy. In mapping, it is always better to go from a finer-grained terminology to a coarser terminology. The second step (SNOMED CT → ICD-11) is responsible for some of the errors in method 1. Since the SI map is generated algorithmically without manual review, some errors are expected.

Failure analysis shows that the main source of error in method 2 is the coverage and accuracy of the WHO map, which goes from ICD-10 to ICD-11. The first step of converting from ICD-10-CM to ICD-10 is usually quite straightforward and seldom results in error. However,

because some ICD-10 codes are not covered by the WHO map, we have to roll-up those ICD-10 codes so that we can use the WHO map to get to an ICD-11 code. This roll-up process reduces accuracy and is the cause of some errors. Errors in the WHO map are another cause of failure in method 2. It seems that some obvious errors (such as the constipation example above) may be due to the map not catching up with the update of ICD-11. But whatever the cause, it is desirable that WHO publishes a high-quality map to facilitate the transition from ICD-10 to ICD-11.

Applications

Even though method 2 is outperformed by method 1, it is still useful. Maps from method 1 can be used to augment and complement those from method 1. This is shown in the combined methods. When we use the union of methods 1 and 2, we get a map that has higher coverage and recall than either method 1 or method 2 used alone. Greater coverage and recall is advantageous, but comes at the expense of precision, which is acceptable from the perspective of assisting manual curation of the map. In practice, the use here is to suggest candidate map targets for mappers to review. Since the candidates will be confirmed or refuted by manual review, the lower precision of the union map will not be a major handicap. On the other hand, if we take the intersection of method 1 and method 2 (i.e., only keeping the map records that are the same in both methods), we will have a map that has higher precision than either method used alone. But this comes at the expense of coverage and recall. The intersection map can be used for tasks that require high precision, such as quality assurance of maps from other sources. In our study, the sequential maps did help us uncover some errors in our reference standard.

Limitations and future work

We recognize the following limitations in our study. The reference standard we used only covers 943 ICD-10-CM codes, a small proportion of all ICD-10-CM codes. However, these are the most commonly used codes based on large data sets. The failure analysis was carried out by one terminologist (JX) and had not been independently corroborated. In the future, we will explore other methods of automatic mapping between ICD-10-CM and ICD-11, including lexical matching and machine learning approaches.

Conclusions

The two sequential mapping approaches between ICD-10-CM and ICD-11, going through SNOMED CT or ICD-10, are useful in automatically creating a draft map from ICD-10-CM to ICD-11 and would reduce manual curation efforts in creating the final map. The various approaches offer different trade-offs among coverage, recall and precision.

References

[1] World Health Assembly Update, 25 May 2019: Available from: https://www.who.int/news-room/detail/25-05-2019-world-health-assembly-update.
[2] Pocai, B., The ICD-11 has been adopted by the World Health Assembly. *World Psychiatry*, 2019. 18(3): p. 371-372.
[3] Editorial, ICD-11: a brave attempt at classifying a new world. Lancet, 2018. 391(10139): p. 2476.
[4] Krawczyk, P., ICD-11 vs. ICD-10 - a review of updates and novelties introduced in the latest version of the WHO International Classification of Diseases. *Psychiatr Danub*. 2020 Autumn-Winter;32(3-4):320-324. doi: 10.24869/psyd.2020.320.
[5] Mortality Data from the National Vital Statistics System. Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/00001356.htm
[6] International Classification of Diseases, (ICD-10-CM/PCS) Transition. Available from: https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm
[7] Fung, K.W., Xu, J., and Bodenreider, O., The new International Classification of Diseases 11th edition: a comparative analysis with ICD-10 and ICD-10-CM. *J Am Med Inform Assoc*, 2020. 27(5): p. 738-746.
[8] Fung, K.W., Xu, J., Synergism between the mapping projects from SNOMED CT to ICD-10 and ICD-10-CM. *AMIA Annu Symp Proc.* 2012;2012:218-27. Epub 2012 Nov 3.
[9] Fung, K.W., Bodenreider, O., et al. Preparing for the ICD-10-CM Transition: Automated Methods for Translating ICD Codes in Clinical Phenotype Definitions. *EGEMS (Wash DC)*. 2016 Apr 12;4(1):1211. doi: 10.13063/2327-9214.1211. eCollection 2016.
[10] Euzenat, J., & Shvaiko, P. (2013). Ontology matching (2nd ed.). Heidelberg (DE): Springer-Verlag.
[11] Zhang, S., Bodenreider O., Experience in Aligning Anatomical Ontologies. *Int J Semant Web Inf Syst*. 2007;3(2):1-26. doi: 10.4018/jswis.2007040101.

Address for correspondence

Email: kfung@mail.nih.gov